Appendix A Additional figures to supplement the methods and results

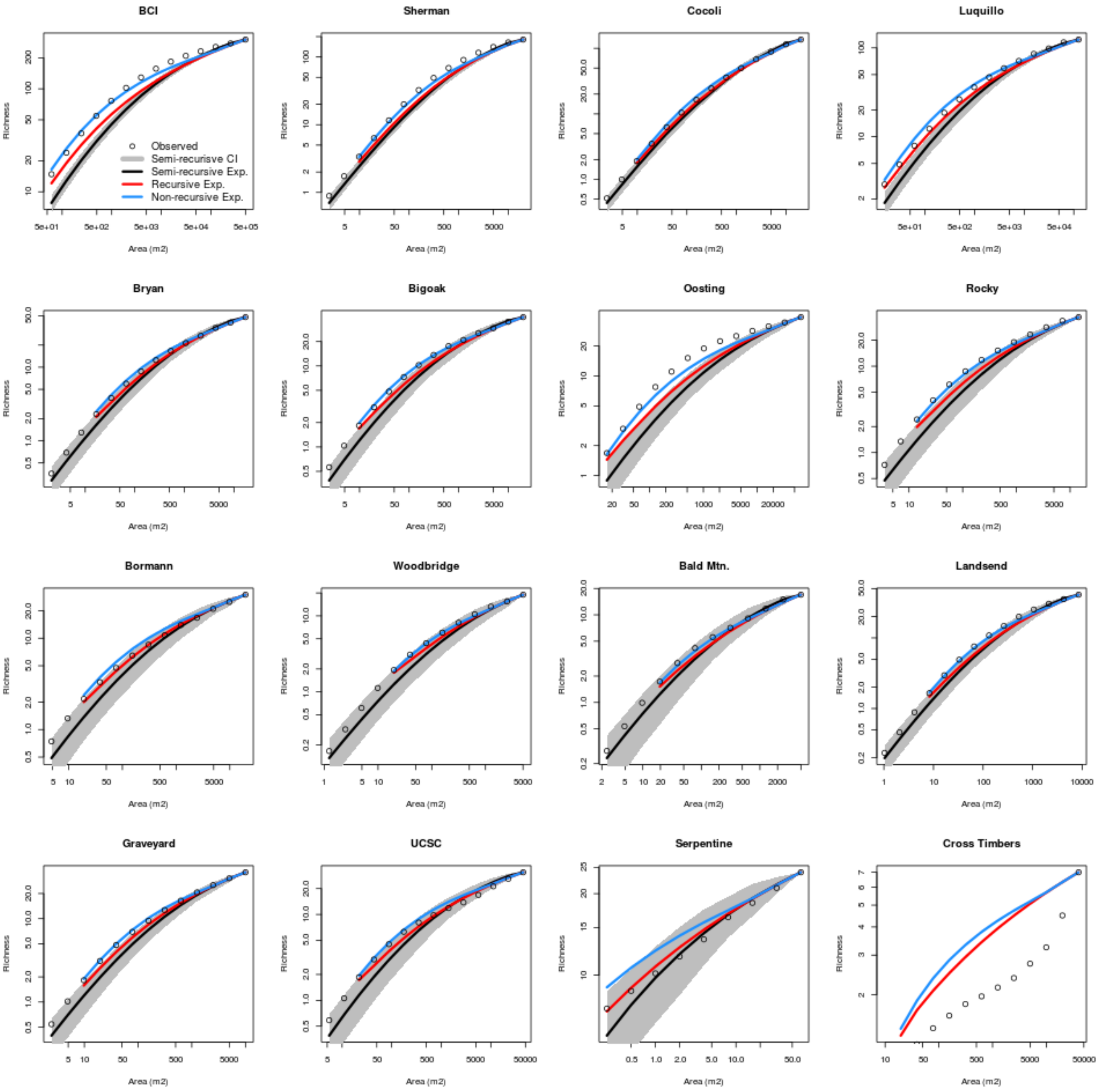

Fig A1. The observed log-log species-area relationships (circles) across the 16 sites relative to the recursive (red line) and non-recursive (blue line) formulations of METE and the simulated semi-recursive METE (black line) when the SAD is inferred from S_0 and N_0. The 95% confidence interval for the semi-recursive prediction is displayed as a grey envelope. A semi-recursive prediction for the Cross Timbers community is not available due to this site's low richness.

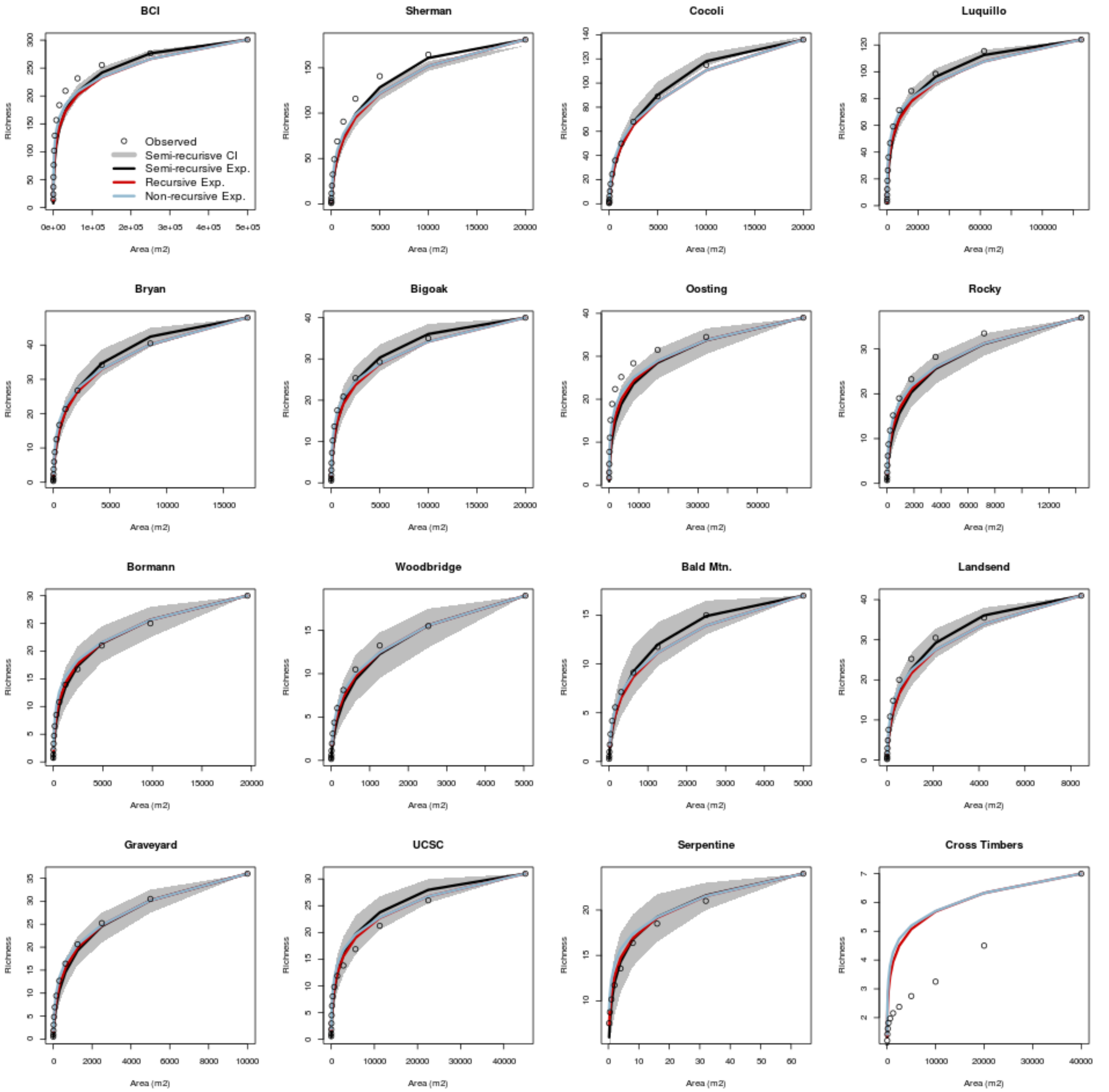

Fig A2. The observed arithmetic species-area relationships (circles) across the 16 sites relative to the recursive (red line) and non-recursive (blue line) formulations of METE and the simulated semi-recursive METE (black line) when the SAD is inferred from S_0 and N_0. The 95% confidence interval for the semi-recursive prediction is displayed as a grey envelope. A semi-recursive prediction for the Cross Timbers community is not available due to this site's low richness.

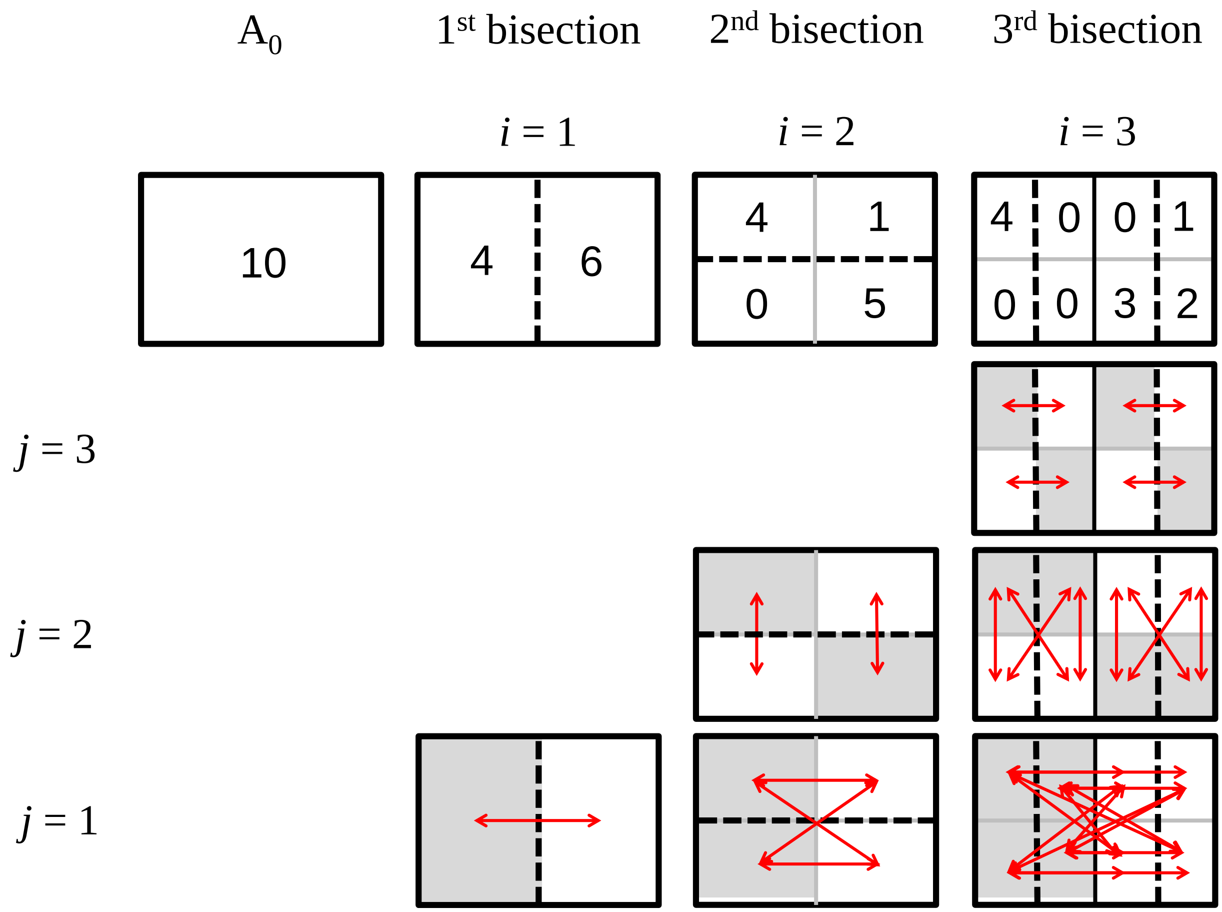

Fig A3. This diagram is an extension of Fig. 1 in the main text and its purpose is to further illustrate the “user rules” of how a landscape can be bisected to different degrees (columns, *i*-values) and how samples are compared for different separation orders (rows, *j*-values). In this specific example, three bisections are used to generate a spatially explicit distribution of 10 individuals. The red arrows illustrate which cells would be compared for a given separation order, *j*. Every possible separation order is illustrated for each level of bisection.

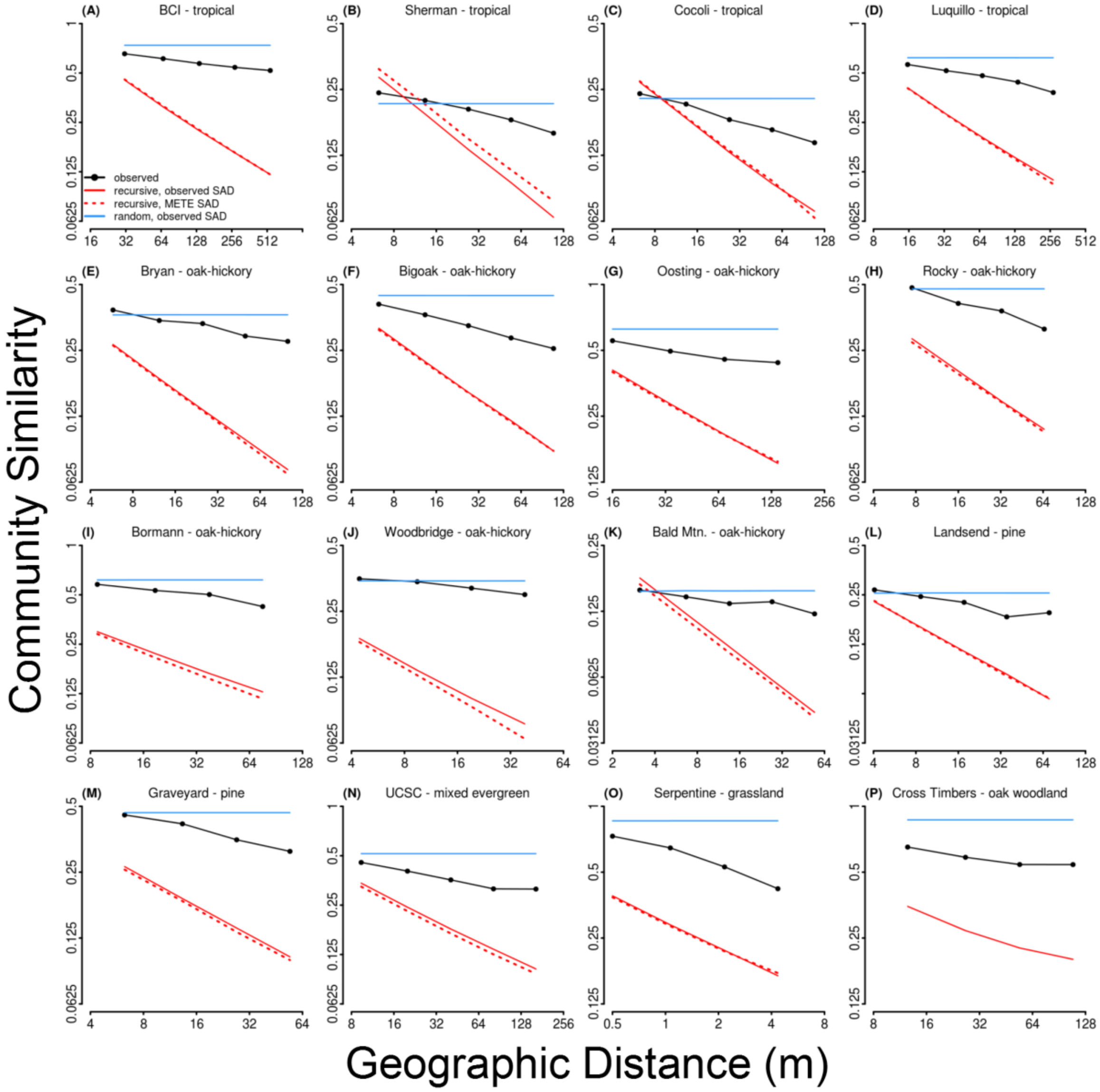

Fig A4. The log-log transformed observed and predicted distance decay relationships for each site at a single spatial grain (identical to Fig. 2 in the main text but with log-log axes). Community similarity represents the average of the abundance-based Sørensen index for each spatial lag. The spatial grain displayed was taken at either 8 or 9 bisections of the total area depending on whether the total extent was a square or a rectangle respectively. Geographic distance was calculated as the average physical distance between the samples compared at given separation order (see Methods and Fig. A3 for additional information).

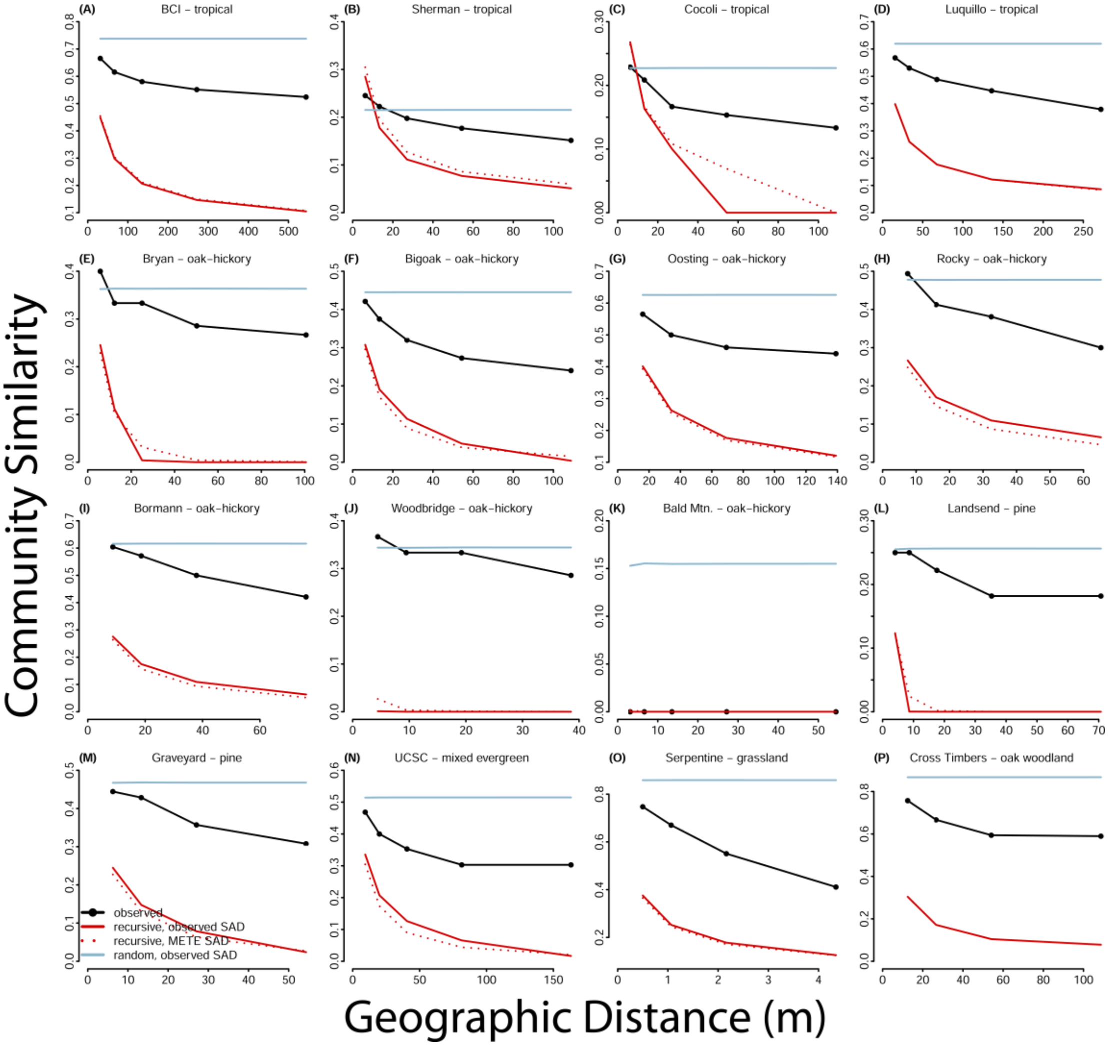

Fig A5. The arithmetic DDR based upon medians (rather than averages) of the abundance-based Sørensen index. The observed (black line with dots) and predicted distance decay relationships (METE: dark grey lines, solid for the observed SAD, dashed for the METE SAD; random placement: light grey line) for each site at a single spatial grain. The spatial grain displayed was taken at either 8 or 9 bisections of the total area depending on whether the total extent of a site was a square or a rectangle respectively. Geographic distance was calculated as the average physical distance between the samples compared at given separation order (see Methods and Fig. A3 for additional information).

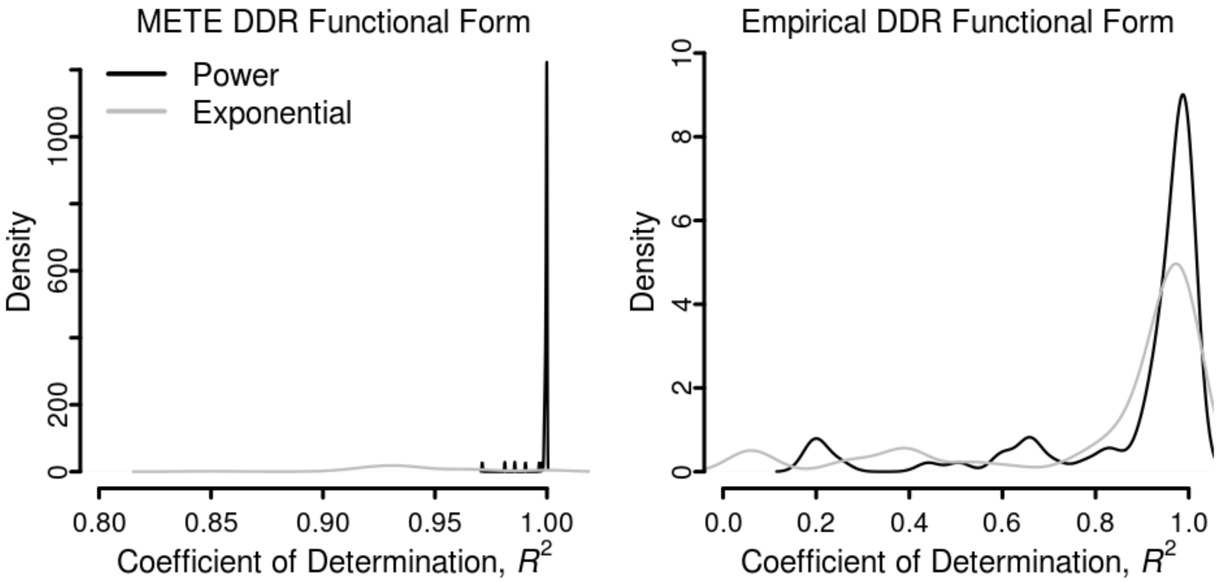

Fig A6. Kernel density lines for the coefficient of determination, R^2, for the fitted power (black line) and exponential models (grey line) for the predicted METE distance decay relationship (DDR, left panel) and for the empirical DDRs (right panel). Both were best fit by a power model of DDR. The R^2-values were computed using weighted least-squares regression where the weights were assigned based on the number of comparisons at a given spatial lag.

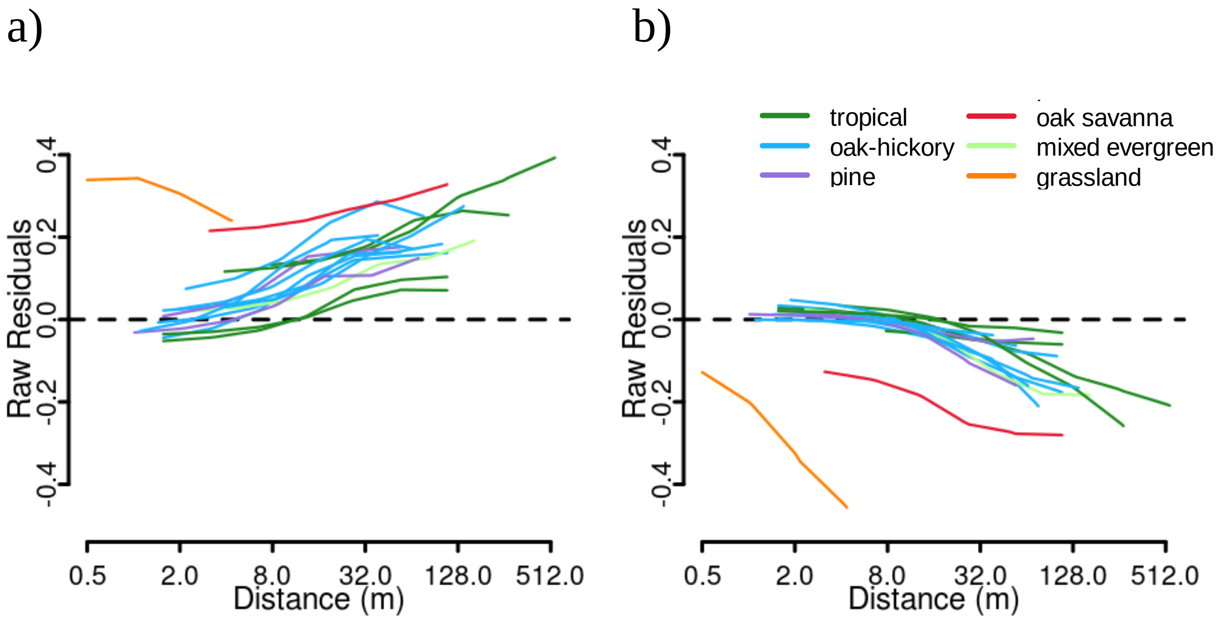

Fig A7. The DDR residuals (i.e., observed – expected) for the empirical datasets relative to the simulated iterative METE (**a**) and RPM (**b**) across all spatial lags and grains. A lowess smoother was applied to the residuals for each site. The datasets are color coded by their habitat type. The dashed line indicates residuals of zero.

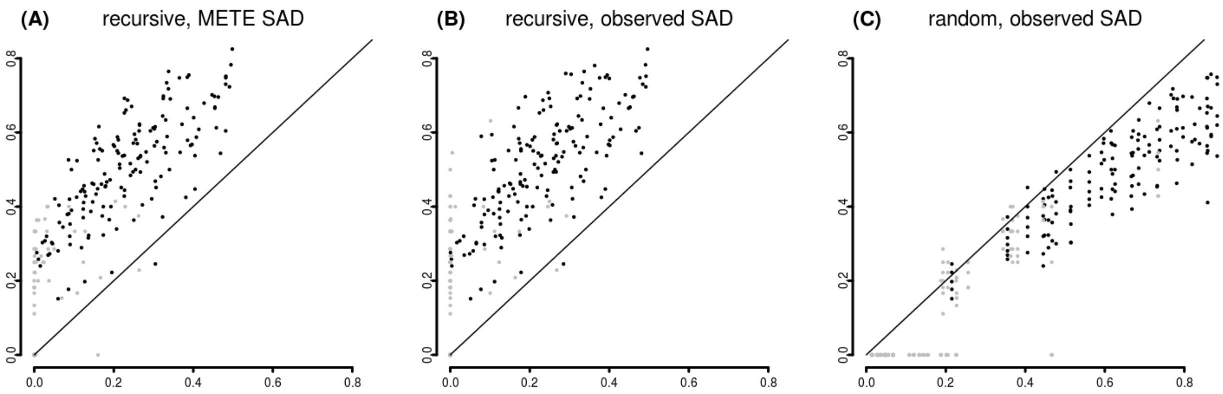

Fig A8. The arithmetic one-to-one plots based on medians (rather than averages) of the predicted and observed abundance-based Sørensen similarity values for the three models across all distances and spatial grains. The solid line is the one-to-one line. The grey points represent values from spatial grains in which the average individual density was low (i.e., less than 10 individuals) and thus both the observed and predicted similarities must be close to zero simply because of a sampling effect.

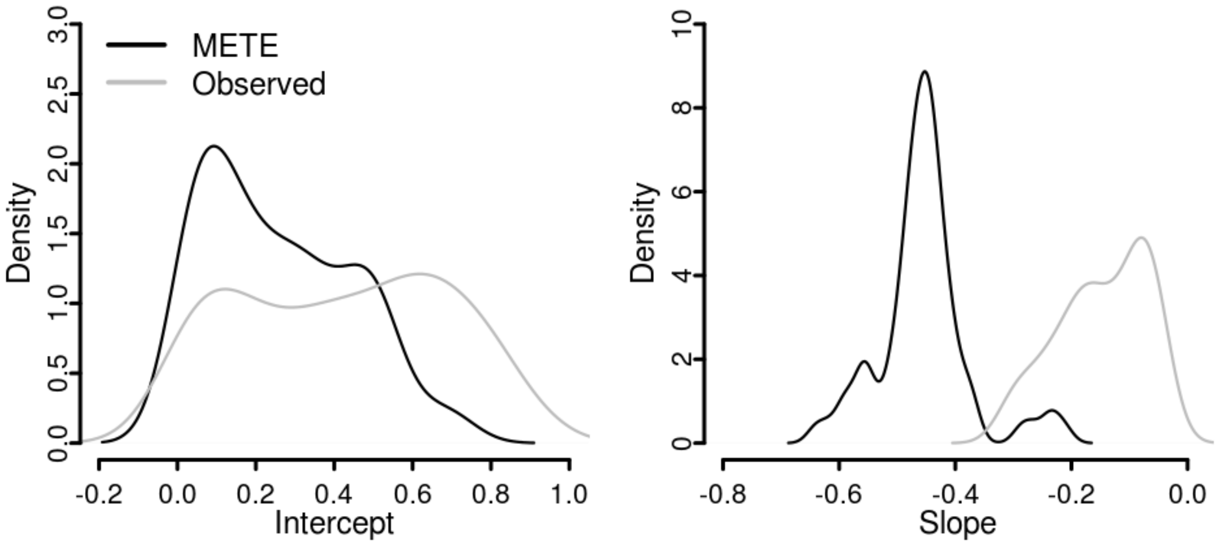

Fig A9. Kernel density plots of the estimated intercept and slope of the power law weighted regression models. The black line illustrates the results for METE with the fixed SAD and the grey line illustrates the observed empirical results. The METE predicted curves had lower intercepts and slopes compared to the empirically observed patterns.

Appendix B

Analytical derivation of community similarity predicted by the semi-recursive formulation of METE and comparison with simulated results

The MaxEnt solution to Π has a simple analytical form for a special case in which the area A is half the total area A_0 (Harte 2011, Eq. 7.51):

$$\Pi(n \mid \frac{A_0}{2}, n_0) = \frac{1}{1+n_0} \qquad \text{(B1)}$$

Equation B1 demonstrates that METE implicitly assumes there is an equal likelihood that any number of individuals between 0 and n_0 will be placed in one of two halves of A_0 or in other words that all spatial arrangements of n_0 individuals between the two cells are equally likely. This prediction is identical to the assumption of the HEAP model (Harte *et al.*, 2005; Harte, 2007). When recursively applying Eq. B1 to successive bisections of A_0 ($A_0/2$, $A_0/4$, …, $A_0/2^i$) the METE prediction for the Π distribution is given by HEAP (Harte, 2011):

$$\Pi(n \mid A_i, n_0) = \sum_{q=n}^{n_0} \frac{\Pi(q \mid A_{i-1}, n_0)}{(q+1)}, \quad i = 1, 2, 3, \ldots \qquad \text{(B2)}$$

where A_i is defined as $A_0/2^i$. Additionally, Harte (2007) derived the equations needed to compute the presence-absence based Sørensen similarity index for HEAP using only the abundance of each species at the anchor scale (A_0). Specifically, Harte (2007) derived the species-level commonality function, $\chi(n_0|A_i, D_j, A_0)$, which is the probability that a species with n_0 individuals is present in two A_i cells separated by the jth bisection but not the j – 1st bisection.

$$\chi(n_0 \mid A_i, D_j) = (n_0+1)^{-1} \sum_{n=2}^{n_0} \chi(n \mid A_{i-1}, D_{j-1}), \quad i = 1, 2, 3, \ldots, \; j = 1, 2, 3, \ldots i \qquad \text{(B3)}$$

and when j = 1,

$$\chi(n_0 \mid A_i, D_1) = (n_0+1)^{-1}\sum_{n=1}^{n_0-1}\left[1-\Pi(0 \mid A_{i-1}, n)\right]\cdot\left[1-\Pi(0 \mid A_{i-1}, n_0-n)\right] \qquad \text{(B4)}$$

Under the METE/HEAP assumption that each species is independently spatially distributed, the community-level commonality function (i.e., the presence-absence Sørensen index) is approximated by:

$$\chi(A_i, D_j) = \sum_{k=1}^{S_0}\chi(n_{0,k} \mid A_i, D_j) \Bigg/ \sum_{k=1}^{S_0} 1-\Pi(0 \mid A_i, n_{0,k}) \qquad \text{(B5)}$$

where $n_{0,k}$ represents the abundance of the kth species at A_0. Eq. B5 is the ratio of the expected number of species in common between two samples and the expected sample richness. It is not an exact formulation for the Sørensen index because it treats the numerator and denominator as two independent quantities (Plotkin & Muller-Landau, 2002). These results are based on using observed species abundances, so to derive a full METE prediction we re-express Eq. B5 in terms of the METE-SAD, $\Phi(n_0 | S_0, N_0)$ at the anchor scale A_0 yielding:

$$\chi(A_i, D_j) = \sum_{n=1}^{N_0}\left[\chi(n \mid A_i, D_j)\cdot\Phi(n \mid S_0, N_0)\right] \Bigg/ \sum_{n=1}^{N_0}\left[\left(1-\Pi(0 \mid A_i, n)\right)\cdot\Phi(n \mid S_0, N_0)\right] \qquad \text{(B6)}$$

To connect METE's spatially explicit predictions with data, distance is treated as a discrete variable that is defined by the geometry of the bisections (Fig 1, Appendix A, Fig A2). Specifically, D_j is the average distance between two A_i cells that are separated by the jth but not the j – 1st bisection where $1 \le j \le i$ (Ostling *et al.*, 2004; Harte, 2007). We will refer to Eqs. B5 and B6 as the semi-recursive versions of METE.

In Fig B1, we compare the analytical and simulated forms of the distance decay relationship (DDR) for the presence-absence Sørensen index as well its two components: the numerator (i.e., average number of species) and the denominator (i.e., shared number of species). This comparison demonstrates that these quantities are closely approximated by Eq. B5;

however, the accuracy of the approximation will decrease as sample area approaches the total area and when there are few individuals on the landscape such that many of the pairwise sample comparisons are undefined (i.e., when the denominator of the Sørensen index is zero).

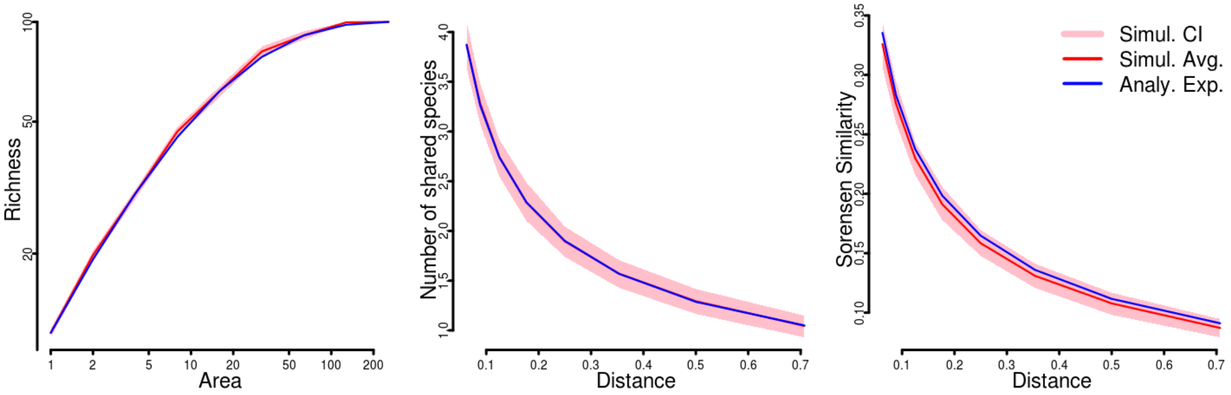

Fig B1. The analytical approximation (Eq. B5, blue line) and the average across 200 simulated communities (red line with pink 95%confidence interval) for the semi-recursive METE species-area relationship (left panel), the distance decay relationship of shared species richness (middle panel), and the distance decay relationship of community similarity (right panel). The input parameters were $S_0 = 100$ and $N_0 = 5000$ with a uniform SAD such that each species had 50 individuals on a square landscape with 256 cells (i.e., 8 bisections).

Literature Cited

Harte, J. (2011) *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics*, Oxford University Press Inc., New York.

Harte, J. (2007) *Toward a mechanistic basis for a unified theory of spatial structure in ecological communities at multiple spatial scales*. *Scaling Biodiversity* Ecological Reviews. (ed. by D. Storch, P.A. Marquet, and J.H. Brown), pp. 101–126. Cambridge University Press, Cambridge.

Harte, J., Conlisk, E., Ostling, A., Green, J.L. & Smith, A.B. (2005) A theory of spatial structure in ecological communities at multiple spatial scales. *Ecological Monographs*, **75**, 179–197.

Ostling, A., Harte, J., Green, J.L. & Kinzig, A.P. (2004) Self-similarity, the power law form of the species-area relationship, and a probability rule: Reply. *American Naturalist*, **163**, 627–633.

Plotkin, J.B. & Muller-Landau, H.C. (2002) Sampling the species composition of a landscape. *Ecology*, **83**, 3344–3356.

Appendix C

Proof that the multi-cell recursive Π distribution is scale consistent

The recursive formulation of the marginal Π distribution for METE is given by (Harte, 2011):

$$\Pi(n \mid A_i, n_0) = \sum_{q=n}^{n_0} \frac{\Pi(q \mid A_{i-1}, n_0)}{(q+1)}, \quad i = 1, 2, 3, ... \tag{C1}$$

Conlisk et al. (2007, Eq. 2.2) provides the multi-cell generalization of C1:

$$P(\mathbf{N} \mid A_i, n_0, c = 2, \phi = 0.5) = \prod_{(n_L, n_R) \in S(\mathbf{N})} \frac{F(a, n_L) F(a, n_R)}{F(2a, n_L + n_R)} \tag{C2}$$

where, $a = (1 - \phi) / \phi = 1$ such that the following simplifications can be made: $F(1, n) = 1$ and $F(2, n) = 1 + n$. Thus C2 simplifies to:

$$P(\mathbf{N} \mid A_i, n_0, c = 2, \phi = 0.5) = \prod_{(n_L, n_R) \in S(\mathbf{N})} \frac{1}{n_L + n_R + 1} \tag{C3}$$

In the above equations, **N** is the spatial configuration of cells across the landscape containing the n_0 individuals, n_L and n_R are the abundances on the left and right side a given bisection, $S(\mathbf{N})$ is the set of bisections leading to landscape **N**, $c = 2$ indicates that a bisection scheme is being used, and $\phi = 0.5$ indicates the aggregation parameter is such that the equal-probability rule predicted by METE is in effect. The key to recognizing that Eq. C3 is scale-consistent (sensu Haegeman & Etienne, 2010) is that it is only defined over $S(\mathbf{N})$. In other words, the sum of the probabilities of the configurations of n_0 individuals at a fine-scale grid with M_1 cells when considered on a coarse-scale grid of M_2 cells is equal to the probability of the configurations solved for directly at M_2 (in the context of bisections, $M_1 = lM_2$ where $l = 2^i$, $i = 1,2,3,\ldots$). To prove that this is generally true we consider n_0 individuals distributed across a landscape that is partitioned into a fine grid with $M_1 = 4$ cells and a coarse grid $M_2 = 2$ cells. Consider the probability that n individuals occur in only the left quadrat of the M_2 cells (i.e., $n_L = n$, $n_R = n_0 - n$

where $n \in \{0, \dots, n_0\}$). There are $n + 1$ configurations (i.e., unique landscapes **N**) of the unlabeled individuals in the M_1 cells that are consistent with n individuals occurring in only the left side of the M_2 grid – the probability of each of these configurations is given by:

$$P(\mathbf{N}) = P(n_L = n, n_R = n_0 - n) \cdot P(n_L = j, n_R = n - j) \tag{C4}$$

Where the first term on the RHS is the probability of observing n individuals on the left hand side of M_2 and the second term on the RHS is the probability for a given distribution of the n individuals across the two M_1 cells with abundance j and $n - j$ within the left M_2 cell. Using Eq. C3 we can express C4 as:

$$P(\mathbf{N}) = \frac{1}{n + (n_0 - n) + 1} \cdot \frac{1}{j + (n - j) + 1} = \frac{1}{(n_0 + 1)(n + 1)} \tag{C5}$$

If we scale-transform this result to compute the probability of observing n individuals only on the left of the M_2 grid we must sum across the probabilities of the $n + 1$ possible configurations which is:

$$\sum_{(n_L, n_R) \in S(\mathbf{N})} P(\mathbf{N}) = \frac{n + 1}{(n_0 + 1)(n + 1)} = (n_0 + 1)^{-1} \tag{C6}$$

Equation C6 is identical to the directly calculated probability of observing n individuals on the left half of the M_2 grid which proves that the Eq. C3 is scale consistent.

Literature Cited

Conlisk, E., Bloxham, M., Conlisk, J., Enquist, B. & Harte, J. (2007) A new class of models of spatial distribution. *Ecological Monographs*, **77**, 269–284.

Haegeman, B. & Etienne, R.S. (2010) Entropy maximization and the spatial distribution of species. *The American Naturalist*, **175**, E74–E90.

Harte, J. (2011) *Maximum Entropy and Ecology: A Theory of Abundance, Distribution, and Energetics*, Oxford University Press Inc., New York.

Appendix D

Code Supplement

We have provided all of the code to replicate the analyses on figshare: http://dx.doi.org/10.6084/m9.figshare.978918. The file mete-spatial.zip should be downloaded and decompressed in what R recognizes as the home directory (the R code will need to be modified if you decide not to use your home directory). R v 2.12 or higher and python 2.6 are needed to run the code. The following python packages are required: matplotlib, mpmath, numpy, and scipy, and the following R packages are required: vegan, hash, RCurl and bigmemory.

It is also possible to access the code via GitHub at the following addresses:

The primary code repository is located at:

https://github.com/weecology/mete-spatial

Additional scripts needed to run the core METE DDR functions are located here:

https://github.com/weecology/METE

https://github.com/weecology/macroecotools

After the files are downloaded (from figshare or GitHub) and decompressed (if using the mete-spatial.zip file) navigate to the METE directory and run the following python command

```
python setup.py install
```

run the same command in the macroecotools directory.

To download the publically available data and run the analysis navigate to the directory ~/mete-spatial and run the command:

```
Rscript ddr_run_all.R
```

This script will download two datasets, analyze them, and graph the results. The plots will appear in the directory ~/mete-spatial/figs/